Distributed Information System for Heterogeneous Environmental Data Integration

Oscar Luiz Monteiro de **Farias**[1], D. Sc. **Margareth Simões** Penello Meirelles[1,2], D.Sc.

[1] Universidade do Estado do Rio de Janeiro
Faculdade de Engenharia
Departamento de Engenharia de Sistemas e Computação
Rua São Francisco Xavier, 524, bloco D, 5[0] andar, sala 5018
– Rio de Janeiro, RJ - Brasil
Tel: (0xx) (21) 587-7442
Fax: (0xx) (21) 587-7374

[2] Embrapa Solos
Empresa Brasileira de Pesquisa em Agropecuária
Centro Nacional de Peqsuisa de Solos
Rua Jardim Botânico, 1024
Rio de Janeiro, RJ - Brasil
Tel: (0xx) (21) 274-4999
Fax: (0xx) (21) 274-9152

ABSTRACT

This work describes the use of a distributed system on an Internet/Intranet environment - DISI^{2}E. This system is being applied to environmental data at the Pantanal area in Brazil, as a tool for accessing several different thematic maps, dealing with different scales, gathering information from multiple sources, and enabling the decision process at municipal and regional level to be interactively accomplished. The system integrates heterogeneous data presented as modules. These modules can be any kind of file visualised by a browser: pictures, html files, audio files, animations files and other types of files visualised via a plug-in, like executable programs (CGI), applets and queries to databases. The user, through an interface very similar to the Windows Explorer, selects these modules. Then, these modules, following the object-oriented paradigm, are related basically in part-of relations (containment relationships, like regions on a map) or some of them – simpler – are taken as attributes of more complex ones. In order to operate DISI^{2}E users need only *http (hyper text transfer protocol) servers* – to publish the distributed information – and *browsers*, enabling users to navigate through the distributed system visualising non-structured information, querying databases and executing programs.

KEYWORDS

Distributed information system, hypertext, object associations, environmental information integration.

1. INTRODUCTION

Agricultural activities have expanded dramatically in the Upper Taquari River basin during the last 20 years, causing significant damage to the original Cerrado ecosystem and threatening the Pantanal biome[1]. The results were degradation of the riparian vegetation and widespread erosion, with river siltation and degradation of the natural resources, risking the sustainability of anthropic activities in the Pantanal lowlands of Mato Grosso do Sul. Pantanal Lowlands is the second most important ecossystem in Brazil, after Amazonia.

To study this process, several municipalities of the region and EMBRAPA joined together in the Taquari Project. The Taquari project aims the development of a Decision Support System (DSS) [2] to assist the land use management of the Upper Taquari Basin, being a useful tool to the environmental decision makers.

In order to attain that goal, the erosion processes and water contamination by agrochemicals are being monitored in three selected sub-basins. The information gathered feed a GIS system [3,4] which, along with Remote Sensing data sources, monitor land use changes and the erosion process throughout the years. This system is part of a decision Support System which applies Artificial Intelligence Techniques to, among others: provide information on the cost/benefit of different mitigation options; rank the municipalities with regards to environmental impact of agricultural activities and rank areas in terms of priority for interventions.

Manipulating all those information the decision maker would: i) identify the more important sources of erosion; ii) disseminate information related to the sources of erosion, between people of the municipalities, in order to create an environmental conscience; iii) disseminate information related to alternative in the land use and in the conservation techniques; iv) aggregate dynamic heterogeneous information. In this context, DISI^{2}E could be used in order to aggregate and disseminate information related to the Taquari project.

DISI^{2}E is a Distributed Information System on an Internet/Intranet Environment. The goal was to build an Internet based distributed system with a graphical user interface based on hierarchical maps, associating information and services to a given topology. In that way, the heterogeneous data generated in Taquari geographical databases could be integrated and queries to answer questions directly related with the environmental aspects could be done, like bio-physical data (soil, geology, geomorphology, vegetation data), social-economical data, erosion data, land use/ land cover data and so on. Besides that, it would be interesting, for example, to show different satellites images or even how to apply conservation techniques in agriculture field - or access maps in different scales of detail, going from the watershed to a micro-watershed located in one critical municipality. Formularies containing social-economical aspects of an area (educational level of the owner, type of conservation practices used, number of cattles, and so on), to be filled and sent to the person associated to administrative tasks, were also desired. Some users also wanted to have information about a neighbour municipality. At the same time, the system could be used to maintain a database related to land use and geographical information dynamically updated by satellite images and having many different cartographic scale and projections.

The system requirements could be thought, in a broader sense, as the possibility of aggregate generic isolated objects that describes environmental aspects [5,6,7] and to relate these objects in a logical manner. Most of the time these are part-of or containment relationships [8,9,10,11]. If we could allow the users to establish arbitrary associations between objects we could solve our problem and, more than this, we would have a system with the possibility of being applied in several distinct areas: commercial, educational (e.g., geography classes), etc. Another important system requirement is that we need a distributed system on an Internet/Intranet environment, i.e., the objects could, at principle, be located at any machine on the Internet address space.

2. LOGICAL SOLUTION FOR DISI^{2}E

In DISI^{2}E we have envisaged a solution very similar to a child's toy called LEGO. With LEGO blocks children build complex objects based on simple ones represented by some elementary blocks. Simple objects are associated on each other, to build more complexes ones. Similarly, we could, for example, think of the natural geographical resources of a region, represented by objects. These objects can be regarded as a set of files of different types, describing physical resources, personal resources, procedures, methodologies, and so on. All we have to do is connect these objects in a logical way to have a model of the geographical region [12,13]. The hypertext technology gives us the possibility to connect these objects together and to establish relations between them in a very general way. Our solution links the objects using hypertext references, but in a more limited way. Basically, we have two kinds of relations between objects: i) a containment relationship, like the ones that we see in maps, where, for example, a country encompasses several states. This can be a recursive (finite) relation, since each state by its turn, encompasses several districts and municipalities; ii) a kind of attribute relationship, where a certain object can have associated descriptors, that give additional information related to the object. For instance, at the Taquari watershed there are 9 municipalities, which are part of thematic maps of the region, such as: (soil map, geological map, geomorphology), thus having lots of files (descriptors). One of these descriptors could be databases, satellite images, pictures showing gullies and landslides and so forth. As long as new information are generated in different level of detail or new satellite images area are aggregated or an aerial photography or a videography of the area is done new descriptors are incorporated.

Our solution also needed to allow some kind of navigation between these maps, and to allow the user to inspect the object descriptors. Then we have represented the objects in a containment relationship as image maps objects, a well-known technique from the World Wide Web (WWW), so that we can use an http browser, like Internet Explorer or Netscape Navigator, to go from one map to another. The descriptors were represented as hyperlinks in the object being browsed, that, when pressed, they call another object describing, or related to, the first one. Descriptors can be any kind of file visualized by a browser: pictures, html files, audio files, animation files and even other types of files visualized via a plug-in (e.g., CGM - Computer Graphics Metafile), like executable programs (CGI), applets, and database queries. The interaction with databases is found in the descriptor. So it can be embedded, via SQL statements, in an applet referenced by a html file, or, in a html file that communicates directly with a http server that hosts PHP. In the former case the database queries are accomplished through JDBC (Java Data Base Connectivity) and, in the latter, through embedded SQL statements in html files.

3. IMPLEMENTATION OF DISI^{2}E

We have developed a software tool called "Developer View", which is somewhat similar, in pictorial terms, to Windows Explorer. The Developer View allows the user to associate the objects - html files, sound files, image files, animation files, SQL queries, CGM files, CGI programs, and other file formats directly interpreted by an http browser - that, together, represent a distributed system. These objects can be located in any directory of the computer system. The associations between these objects are built through a graphical interface developed in C++ Builder [14]. They are restricted to the two types of relations mentioned above: containment and attribute. The net result of all these associations is a web site, written in html. All files that compose the web site: the original objects given by the user, plus the association between these objects - reflected in the html files - are put in an htdocs directory of an http server. Finally, the user points his browser to the root object of the htdocs directory on the http server and begins his navigation through a true distributed system on an Internet/Intranet environment.

When creating the site, the Developer View automatically builds all the directory structure needed, obeying the paths defined by the user. Some very useful functions help the developer task. They are: *i) Autosave* - when enabled, the Developer View saves the work done by the developer in building the site at previously defined time intervals; *ii) Lock/Unlock objects* - allows the developer to maintain some control over the objects that are already set to be published and the objects still in development. Locked objects cannot be edited; *iii) Dependency Verification* - sweep all the structure of the site that is being generated, seeking for duplicated objects, which are not allowed, and for invalid associations. All the inconsistencies found are registered in a log.

The Developer View permits the fast development of sites, and also reduces the effort to maintain sites, since it comprises site edition.

4. TOOLS NEEDED BY THE USER

All the user needs to operate DISI^{2}E is *http (hyper text transfer protocol) servers* – to disseminate the distributed information – and *browsers*, so that people can navigate through the distributed system (built by the user at his will) visualizing non-structured information, querying databases and executing programs. Of course, the user needs to have at hand all the objects that he wants to take part on his distributed system and also the resources to make these objects functional. So, if he wants to access a database and submit queries, he will need a Data Base Management System (DBMS), a database and a mechanism to query the database (e.g., an applet using JDBC). The same is true for objects of other types, like CGM animations, sound files, etc.

5. CONCLUSIONS

Based in DISI^{2}E a user can build a distributed system in an incremental way, adding to the systems all the objects that can be normally visualized by a browser. New types of files, not yet invented, will be naturally added to the system, since we hope that the browsers will be capable of dealing with them. The system implicitly uses all the Internet standards, what gives it great flexibility. It is a true multi-platform distributed system, and is prepared to cope with future progress on the Internet. The system can solve the problem of aggregating distributed heterogeneous environmental information, integrating many different type of thematic maps, satellite images, pictures, graphics, social-economical tables, databases and GIS system, as objects being a tool for the decision making process at municipal and regional levels. The goal of easy-access information integration was achieved, as the system involves only simple internet navigation, allowing the operation even for not skilled people, such as administrative clerks of the municipalities. This is a way of not only gathering all kind of environmental data of a region but also to popularise access to those information.

6. REFERENCES

[1] BRASIL. Ministério do Meio Ambiente, dos Recursos Hídricos e da Amazonia Legal. Secretaria de Coordenação dos Assuntos de Meio Ambiente. Programa Nacional do Meio Ambiente. Projeto Pantanal. Plano de Conservação da Bacia do Alto Paraguai: PCBAP. Brasília, 1997. v.3: Análise Integrada e Prognóstico da Bacia do Alto Paraguai.

[2] ZHU, X.; Healey, R. "Towards intelligent spatial decision support integrating Geographic Information Systems and Expert Systems."; GIS/LIS Proceedings; vol. 2, 1992. pp. 877 - 886.

[3] BURROUGH, P.A. Principles of geographical information systems for land resources assessment. Oxford: Clarendon Press, 1996. 194p.

[4] COHEN, D.J.; SHIRLEY, W.L. Integrated planning of information systems.In: MAGUIRE, D.J.; GOODCHILD, M.F.; RHIND, D.W., ed. Geographical Information Systems. Essex: Longman Scientific & Technical, 1991. v.2, p.297-310.

[5] WEBSTER, C.J. and Omare, C.N. “A Formal Approach to Object-oriented Spatial Database Design.”; Proceedings of the European Conference on GIS; EGIS Foundation, Utrecht, 1991. pp 1209 - 1218.

[6] WORBOYS, M.F. Hearnshaw, H.M., Maguire, D.J., "Object-oriented data modeling for spatial databases.", Int. Journal of Geographical Information Systems, 1990, vol. 4, nº. 4. pp. 369 - 383.

[7] WORBOYS, M. F. “GIS – A Computing Perspective.”; Taylor & Francis; London, 1995; pp. 376.

[8] MARTIN FOWLER & KENDALL SCOTT. UML Distiled. Addison-Wesley Object Technology Series, 1997

[9] JAMES J. ODELL. Advanced Object-Oriented Analysis & Design Using UML. Cambridge University Press, 1998.

[10] IVAR JACOBSON, GRADY BOOCH & JAMES RUMBAUGH. The Unified Software Development Process. Addison Wesley, 1999.

[11] ERICH GAMMA, RICHARD HELM, RALPH JOHNSON & JOHN VLISSIDES. Design Patterns – Elements of Reusable Object-Oriented Software. Addison Wesley, 1995.

[12] AYBET, J. “Interoperability of GIS in a Open System Environment.”. Proceedings of the Joint European Conference and Exhibition on Geographical Information; IOS Press; Amsterdam; Volume I; 1996. pp.63-72;

[13] GLOVER, J. “The Integration and Interoperability Challenge: The Need for OpenGIS.”; in Green, D.R.; Rix, D.; Corbin, C. “The AGI source book for GIS - 1996.”; Taylor & Francis; London; 1996.

[14] CHARLIE CALVERT. C++ Builder 3 Unleashed, SAMS Borland Press.